

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 22:29:57 ; Search time 2180 Seconds
(without alignments)
7013.092 Million cell updates/sec

Title: US-09-701-023-1_COPY_81_1024
Perfect score: 944
Sequence: 1 tgcgcactctctctcttc.....atttgtcttgaagctttaa 944

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estl:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_estfun:*
- 15: em_estom:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	312.4	33.1	592	17 B98482	B98482 T25H20TRB T
C 2	296.8	31.4	508	17 B67977	B67977 T25H20TR TA
C 3	182.2	19.3	641	17 BH472044	BH472044 BOHUA43TR
C 4	59.4	6.3	775	17 BH472100	BH472100 BOHUR55TR
C 5	45	4.8	645	10 BE188445	BE188445 pVb 292 L
C 6	41	4.3	1672	10 BE622474	BE622474 601441274

C 7	40.8	4.3	937	17 CNS006XP	AL066056 Drosophila
C 8	40.6	4.3	863	17 BH154230	BH154230 ENTR245TF
C 9	40.4	4.3	552	17 AQ338133	AQ338133 HS_5011_B
C 10	39	4.1	859	17 CNS06CGK	AL394490 T3 end of
C 11	39	4.1	519	14 BQ515341	BQ515341 EST622756
C 12	39	4.1	522	13 B1432782	B1432782 EST535543
C 13	39	4.1	525	14 BQ515340	BQ515340 EST622755
C 14	39	4.1	1105	17 CNS07ARO	AL436930 T3 end of
C 15	39	4.1	1167	12 BG856706	BG856706 1024047G0
C 16	38.8	4.1	638	14 BQ390257	BQ390257 NISC_mq12
C 17	38.8	4.1	673	9 AL647711	AL647711 AL647711
C 18	38.6	4.1	937	17 CNS006ST	AL065880 Drosophila
C 19	38.4	4.1	637	17 AZ097834	AZ097834 RPCI-23-4
C 20	38.4	4.1	653	13 B1927000	B1927000 EST545889
C 21	38.2	4.0	477	13 B1465229	B1465229 1e23h04 Y
C 22	38	4.0	487	12 BF097623	BF097623 EST415696
C 23	37.8	4.0	565	17 AQ490219	AQ490219 RPCI-11-2
C 24	37.4	4.0	331	12 BF587816	BF587816 FMJ_33_H1
C 25	37.4	4.0	921	17 CNS04NCO	AL298521 Tetradon
C 26	37.2	3.9	544	9 AL630678	AL630678 AL630678
C 27	37.2	3.9	551	14 BQ418553	BQ418553 1k46b06 Y
C 28	37.2	3.9	889	17 B21212	B21212 T32J11-sp6
C 29	37.2	3.9	997	17 CNS005TE	AL060767 Drosophila
C 30	37	3.9	660	17 AG142495	AG142495 Pan trogl
C 31	37	3.9	695	17 AZ938420	AZ938420 2M0197C06
C 32	36.8	3.9	275	17 AZ768950	AZ768950 1M0569M08
C 33	36.8	3.9	609	13 B1514873	B1514873 B160016B
C 34	36.8	3.9	636	10 BB635500	BB635500 BB635500
C 35	36.8	3.9	902	13 B1856118	B1856118 603382821
C 36	36.6	3.9	443	17 AQ850356	AQ850356 LMAJFV1_1
C 37	36.6	3.9	550	17 AQ847178	AQ847178 T1378205
C 38	36.6	3.9	609	17 AQ847422	AQ847422 LMAJFV1_1
C 39	36.6	3.9	688	17 AZ775117	AZ775117 2M0007108
C 40	36.6	3.9	1578	12 BF575807	BF575807 602134848
C 41	36.4	3.9	238	17 AZ263790	AZ263790 RPCI-23-4
C 42	36.4	3.9	374	12 BE762203	BE762203 NXCI-083
C 43	36.4	3.9	503	17 BH077388	BH077388 RPCI-24-3
C 44	36.4	3.9	665	17 AQ630761	AQ630761 RPCI-11-4
C 45	36.2	3.8	684	17 AG040879	AG040879 Pan trogl

ALIGNMENTS

RESULT 1
B98482/c

LOCUS
DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B98482 T25H20TRB TAMU Arabidopsis thaliana genomic clone T25H20, DNA sequence.

B98482

B98482.1 GI:3000561

GSS.

Arabidopsis thaliana

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 592)

Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Adams, M.D. and Venter, J.C.

A BAC End Sequence Database for Identifying Minimal Overlaps in

Arabidopsis Genomic Sequencing. Update 3

Unpublished (1997)

Other_GSSs: T25H20TF T25H20TRB T25H20TR

Contact: Steve Rounsley

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: rounsley@tigr.org

Seq primer: M13 Reverse

Class: BAC ends
High quality sequence stop: 592.

FEATURES

Location/Qualifiers

1..592
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T25H20"
/clone.lib="TAMU"
/sex="hermaphrodite"
/note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"

BASE COUNT 195 a 127 c 91 g 176 t 3 others

ORIGIN

Query Match 33.1%; Score 312.4; DB 17; Length 592;
Best Local Similarity 82.1%; Pred. No. 9.9e-82;
Matches 412; Conservative 0; Mismatches 1; Indels 89; Gaps 1;
529 AGAAGAAACGTTGGATGGTGATCAGATAATAGTTCGATCCCAACGGTGGTGATTTT 588
|||||
503 AGAAGAAACGTTGGATGGTGATCAGATAATAGTTCGATCCCAACGGTGGTGATTTT 444
|||||
509 CGAATACACAAATGATTCCTCTCCGATGAACGGCTAGCAGATCTTCTTCAATCAG 648
|||||
443 CGAATACACAAATGATTCCTCTCCGATGAACGGCTAGCAGATCTTCTTCAATCAG 384
|||||
649 ATCATCATCAGAGGACCAAGGTTTCCTTTATGATCATAGATCGCTAGACAGCTTCAG 708
|||||
383 ATCATCATCAGAGGACCAAGGTTTCCTTTATGATCATAGATCGCTAGACAGCTTCAG 324
|||||
709 TTTCTGCTTCTAGTACTACTATTAATCTCTTATTCACGAGGCAACAATCATATC 763
|||||
323 TTTCTGCTTCTAGTACTACTATTAATCTCTTATTCACGAGGCAACAATCATATC 264
|||||
764 ----- 763
263 TAAGTATAGTCCATTTATTAATCTCATATATAGTATATATATATATATATATAT 204
|||||
764 -----GGACCAATGGAGGAATTTGGAGCTACATGGAAGG 799
|||||
203 TATTTGATTTAACTGCTGGGTTTAGGACCAATGGAGGAATTTGGAGCTACATGGAAGG 144
|||||
800 AAACCCTAGAATGGATCAGAGGTGTGAAGGATGACAGTTTTTCCGGGAAATATGG 859
|||||
143 AAACCCTAGAATGGATCAGAGGTGTGAAGGATGACAGTTTTTCCGGGAAATATGG 84
|||||
860 TGAAGAGTTTCACTGGTGCTACAACTGCTCACTCTGATGATGATGATGATGATGATGAT 919
|||||
83 TGAAGAGTTTCACTGGTGCTAAACGCTGCTCACTCTGATGATGATGATGATGATGATGAT 24
|||||
920 CATTGATTTGCTTGAAGCTT 941
|||||
23 CATTGATTTGCTTGAAGCTT 2

RESULT 2 B67977 508 bp DNA linear GSS 09-DEC-1997
LOCUS T25H20TR TAMU Arabidopsis thaliana genomic clone T25H20, DNA
DEFINITION sequence.
ACCESSION B67977
VERSION B67977.1 GI:2666731
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 508)
Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
,J.C.

TITLE

Arabidopsis Genomic Sequencing. Update 3
Unpublished (1997)
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 508.

JOURNAL

COMMENT

FEATURES

source

1..508
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T25H20"
/clone.lib="TAMU"
/sex="hermaphrodite"
/note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"

BASE COUNT 158 a 117 c 82 g 151 t

ORIGIN

Query Match 31.4%; Score 296.8; DB 17; Length 508;
Best Local Similarity 81.6%; Pred. No. 4.1e-77;
Matches 408; Conservative 0; Mismatches 2; Indels 90; Gaps 2;
531 AAGAAACGTTGGATGGTGATCAGATAATAGTTCGATCCCAACGGTGGTGATTTTCG 590
|||||
508 AAGAAACGTTGGATGGTGATCAGATAATAGTTCGATCCCAACGGTGGTGATTTTCG 449
|||||
591 AAATACAAATGATTCCTCTCCGATGAACGGCTACGATCATATCTTCTTCAATCAGAT 650
|||||
448 AAATACAAATGATTCCTCTCCGATGAACGGCTACGATCATATCTTCTTCAATCAGAT 389
|||||
651 CATCATCAGAGGACCAAGGTTTCTTATGATCATAGATCGCTAGACAGCTTCAGTT 710
|||||
388 CATCATCAGAGGACCAAGGTTTCTTATGATCATAGATCGCTAGACAGCTTCAGTT 329
|||||
711 TCTGCTTCTAGTACTACTATTAATCTTATTCACGAGGCAACAATCATATC 763
|||||
328 TCTGCTTCTAGTACTACTATTAATCTTATTCACGAGGCAACAATCATATC 269
|||||
764 ----- 763
268 AGTATATCTCCATTTATTAATCTCATATATAGTATATATATATATATATATATAT 209
|||||
764 -----GGGACCAATGGAGGAATTTGGAGCTACATGGAAGAA 801
|||||
208 TTTGATTTAACTGGTGGGTTTAGGACCAATGGAGGAATTTGGAGCTACATGGAAGAA 149
|||||
802 ACCCTAGAAATGGATCAGAGGTGTGAAGGATGACAGTTTTCGCGGGAATATAGGT 860
|||||
148 ACCCTAGAAATGGATCAGAGGTGTGAAGGATGACAGTTTTCGCGGGAATATAGGT 89
|||||
861 GAAAGAGTTTTCAGTGGGTGCTCAACGCTCCTACTCTAGTGTGATTCAGTCTTAATACC 920
|||||
88 GAAAGAGTTTTCAGTGGGTGCTCAACGCTCCTACTCTAGTGTGATTCAGTCTTAATACC 29
|||||
921 ATTGATTTGTCCTTGAAGCT 940
|||||
28 ATTGATTTGTCCTTGAAGCT 9

RESULT 3

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

DEFINITION

ACCESSION

LOCUS

Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATSLFMSTDSQNSVGNPNDLRLNRLVNVSSGGEIRTTETLKSRRKPGSKTGQOKKQKPT 60
 DB 1 MATSLFMSTDSQNSVGNPNDLRLNRLVNVSSGGEIRTTETLKSRRKPGSKTGQOKKQKPT 60

QY 61 LRGMGVAKLERQRIEERKKQLAAATVGDTSVASISNNATRLPVPDPGVVLOGFPSSLG 120
 DB 61 LRGMGVAKLERQRIEERKKQLAAATVGDTSVASISNNATRLPVPDPGVVLOGFPSSLG 120

QY 121 SNRIYCGVGSGQWIMDPVISPWGFVETSTTHLSLSSINPOMFNASSNNRCDTCFKKKR 180
 DB 121 SNRIYCGVGSGQWIMDPVISPWGFVETSTTHLSLSSINPOMFNASSNNRCDTCFKKKR 180

QY 181 LDGQNNVVRNSGGGFSKYTIMPPMNGYDQYLQSDHHQHSQGLYDHRITARAASVSAS 240
 DB 181 LDGQNNVVRNSGGGFSKYTIMPPMNGYDQYLQSDHHQHSQGLYDHRITARAASVSAS 240

QY 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGKEYEFPFGKYGERSVWATSSLV 300
 DB 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGKEYEFPFGKYGERSVWATSSLV 300

QY 301 GDGCSPTNTIDLSLKL 314
 DB 301 GDGCSPTNTIDLSLKL 314

RESULT 2
 081836 PRELIMINARY; PRT; 314 AA.

AC 081836;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Hypothetical 34.1 kDa protein (NOZ4LE/SPOROCYTELESS).
 GN MA122.140 OR A74G27330.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Reichert B.J., Barel E., Hoheisel J., Mewes H.W., Mayer K.,
 RA Schueller C., Bevan M.;
 RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 SEQUENCE FROM N.A.
 RA Reichert B.J., Barel E., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RN RN
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL030978; CAA19727.1; -;
 DR EMBL: AL161566; CAB79588.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 314 AA; 34078 MW; B0236D910544CEC3 CRC64;

Query Match 99.6%; Score 1628; DB 10; Length 314;
 Best Local Similarity 99.7%; Pred. No. 3.5e-130;
 Matches 313; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MATSLFMSTDSQNSVGNPNDLRLNRLVNVSSGGEIRTTETLKSRRKPGSKTGQOKKQKPT 60
 DB 1 MATSLFMSTDSQNSVGNPNDLRLNRLVNVSSGGEIRTTETLKSRRKPGSKTGQOKKQKPT 60

QY 61 LRGMGVAKLERQRIEERKKQLAAATVGDTSVASISNNATRLPVPDPGVVLOGFPSSLG 120
 DB 61 LRGMGVAKLERQRIEERKKQLAAATVGDTSVASISNNATRLPVPDPGVVLOGFPSSLG 120

QY 121 SNRIYCGVGSGQWIMDPVISPWGFVETSTTHLSLSSINPOMFNASSNNRCDTCFKKKR 180
 DB 121 SNRIYCGVGSGQWIMDPVISPWGFVETSTTHLSLSSINPOMFNASSNNRCDTCFKKKR 180



sporogenesis and encodes a novel nuclear protein
 Genes Dev. 13 (16), 2108-2117 (1999)
 99396716
 JOURNAL MEDLINE
 10465788
 PUBLISHED
 2 (bases 1 to 1302)
 REFERENCE
 Yang, W.-C., Ye, D., Xu, J. and Sundaresan, V.
 Direct Submission
 TITLE
 Submitted (14-JUN-1999) Plant Molecular Genetics, Institute of
 Molecular Agrobiology, 1 Research Link, Singapore 117604, Singapore
 Location/Qualifiers
 1. 1302
 /organism="Arabidopsis thaliana"
 /cultivar="Landsberg erecta"
 /db_xref="taxon:3702"
 /chromosome="4"
 /map="Pc11-mil23"
 /note="nucleus"
 1. 1302
 /gene="SPL"
 80. 1024
 /gene="SPL"
 /note="MADS-box related protein"
 /codon_start=1
 /product="sporocytelless"
 /protein_id="A045344.1"
 /db_xref="GI:5566240"
 /translation="MATSLEFMTDQNSVGNPNLRLNRLVNSGEIRTELKSRG
 RKFGSTGQOKKPTLRGMVAKLERIEKKKOLAAATVGDTSVASISNNATRL
 PVPDGVQLQGFPSLSGSKRIYCGVSGOVMDPVI SPGVEVETSTTHELSSISN
 POMFNASNNRCDTCFKKRLDGDONNVRSNGGFSKYTWI PPMNGYDQYLLQSDH
 HORSQFLYDHRTHARASVSASSTIINPYNEATNHTGPMEEFGSYMEGNPRNGSGV
 KEYEFPKKGIRGVSVVATSSVLVDCSPNTIDLKSL"
 BASE COUNT 402 a 273 c 275 g 352 t
 ORIGIN
 Query Match 100.0%; Score 944; DB 8; Length 1302;
 Best Local Similarity 100.0%; Pred. No. 2.8e-260;
 Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGGCGACTCTCTCTTCTTCAATGATCAACAGATCAAAATCCGTCGGAAACCCAAACGATC 60
 Db 81 TGGCGACTCTCTCTTCTTCAATGATCAACAGATCAAAATCCGTCGGAAACCCAAACGATC 140
 QY 61 TTCTGAGAACACCCGCTTCTCTCAATAGCTCCGGGAGATCCGACAGACACTGA 120
 Db 141 TTCTGAGAACACCCGCTTCTCTCAATAGCTCCGGGAGATCCGACAGACACTGA 200
 121 AGATCTGCTGGTCCGAAACCCAGGATCGAAGACAGCTCAGCAAAACAGAGAAACCAACGT 180
 201 AGATCTGCTGGTCCGAAACCCAGGATCGAAGACAGCTCAGCAAAACAGAGAAACCAACGT 260
 QY 181 TGAGAGAAATGGGTGTAGCAAGCTCGAGCTCAGAGAAATCGAAGAAAGAAAGCAAC 240
 Db 261 TGAGAGAAATGGGTGTAGCAAGCTCGAGCTCAGAGAAATCGAAGAAAGAAAGCAAC 320
 QY 241 TCGCCGCCGCCACAGTCGGAGACAGCTCATAGTAGATCGATCTCTAACAACGCTACCC 300
 Db 321 TCGCCGCCGCCACAGTCGGAGACAGCTCATAGTAGATCGATCTCTAACAACGCTACCC 380
 QY 301 GTTTACCCGTACCGGTAGACCCGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGA 360
 Db 381 GTTTACCCGTACCGGTAGACCCGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGA 440
 QY 361 GCAACAGGATCTATTGTGGTGGAGTCGGGTCCGATGATGATGATGATGATGATGATGATGAT 420
 Db 441 GCAACAGGATCTATTGTGGTGGAGTCGGGTCCGATGATGATGATGATGATGATGATGATGAT 500
 QY 421 CTCATGGGGTTTGTGTGAGACCTCTCCACTACTCATGAGCTCTCTTCAATCTCAATC 480
 Db 501 CTCATGGGGTTTGTGTGAGACCTCTCCACTACTCATGAGCTCTCTTCAATCTCAATC 560
 QY 481 CTCAATGTTTAAAGCTTCTTCCAAATATCGCTGTCGACACTTCTTCAAGAAAGACGTT 540

Db 561 CTCAAAATGTTTAAACGCTTCTTCCAAATATCGCTGTGACACTTGCCTCAGAAAGAACGTT 620
 QY 541 TGGATGGTGCATCAGAAATATAGTAGTTCGATCAACCGGTGGTGGATTTTCGAAATACACAA 600
 Db 621 TGGATGGTGCATCAGAAATATAGTAGTTCGATCAACCGGTGGTGGATTTTCGAAATACACAA 680
 QY 601 TGATTCCTCTCCGATGAACGGCTAGCATGATCTTCTTCAATCAGATCATCATCAGA 660
 Db 681 TGATTCCTCTCCGATGAACGGCTAGCATGATCTTCTTCAATCAGATCATCATCAGA 740
 QY 661 GGAGCAAGGTTTCTTATGATCATAGAAATCGCTAGACAGCTTCAGTTTCTGCTTCTA 720
 Db 741 GGAGCAAGGTTTCTTATGATCATAGAAATCGCTAGACAGCTTCAGTTTCTGCTTCTA 800
 QY 721 GTACTACTATTATTCCTTATTTCAACGAGGACAAATATACAGGACCAATGGAGGAAT 780
 Db 801 GTACTACTATTATTCCTTATTTCAACGAGGACAAATATACAGGACCAATGGAGGAAT 860
 QY 781 TTGGAGCTACATGAAGAAACCTAGAAATGGAATGAGGAGGTGTAAGGAGTACGAGT 840
 Db 861 TTGGAGCTACATGAAGAAACCTAGAAATGGAATGAGGAGGTGTAAGGAGTACGAGT 920
 QY 841 TTTTCCGGGAAATATGTTGAAAGAGTTTCAGTGGTGGCTACAACTGCTCAGTCTAG 900
 Db 921 TTTTCCGGGAAATATGTTGAAAGAGTTTCAGTGGTGGCTACAACTGCTCAGTCTAG 980
 QY 901 GTGATTGAGTCCCTAATACCATGATTTGCTTCTGTAAGACTTTAA 944
 Db 981 GTGATTGAGTCCCTAATACCATGATTTGCTTCTGTAAGACTTTAA 1024
 RESULT 2
 AFI46794
 LOCUS AFI46794 1133 bp DNA linear PLN 29-SEP-1999
 DEFINITION Arabidopsis thaliana NOZZLE (N22) gene, complete cds.
 ACCESSION AFI46794
 VERSION AFI46794.1 GI:5007008
 KEYWORDS
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 1133)
 REFERENCE Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevallier, D.,
 Wisman, E. and Schneitz, K.
 TITLE Molecular analysis of NOZZLE, a gene involved in pattern formation
 and early sporogenesis during sex organ development in Arabidopsis
 thaliana
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11664-11669 (1999)
 MEDLINE 99432290
 PUBLISHED 10500234
 REFERENCE 2 (bases 1 to 1133)
 AUTHORS Schiefthaler, U., Balasubramanian, S., Chevallier, D., Sieber, P. and
 Schneitz, K.
 TITLE Direct Submission
 JOURNAL Submitted (29-APR-1999) Institute of Plant Biology, University of
 Zurich, Zollikerstr. 107, Zurich, ZH 8008, Switzerland
 FEATURES
 Location/Qualifiers
 1. 1133
 /organism="Arabidopsis thaliana"
 /cultivar="Landsberg erecta"
 /db_xref="taxon:3702"
 /map="between PGI1 and PRHA"
 /chromosome="4"
 <1..>1133
 /gene="N22"
 join(<1..531,631..864,954..>1133)
 /gene="N22"
 /product="NOZZLE"
 join(1..531,631..864,954..1133)
 /gene="N22"
 /note="Involved in ovule and anther development and early

sporogenesis"
/codon_start=1
/product="NOZZLE"
/protein_id="AA037775.1"
/db_xref="GI:5007009"
/translation="MATSLFMTSDQNSVGNPNLDLNRNLVNSSEIRTEITLKS
RPGSTGQKQKPLRLGMVAKLERIEKKQLAAATVGTSSVASISNNATRL
PVPDGVVLQCFPSLGSNRIYCGVSGQVMDIPVISPWFVETSTTHLSISIN
POMFNASSNNRCPTCFKKRLDGDONNVRNSNGGFSKYTWIPPMNGYDQYLLOSDH
HORSOCFLYDHRITARAASVSASSTINPYNENATNHTNPMEEFGSYMEGNPRNSGGV
KEYEFPFGYGERVSVAATSSSLVGDSPNTIDLSKL"
variation 90

/gene="NZ2"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="c"
1072
variation 1072

/gene="NZ2"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="a"
323 a 231 c 255 g 324 t

ORIGIN

Query Match 78.0%; Score 736; DB 8; Length 1133;
Best Local Similarity 83.4%; Pred. No. 1.9e-200;
Matches 944; Conservative 0; Mismatches 0; Indels 188; Gaps 2;

QY 1 TGGCGACTCTCTCTCTCTCATGTCACAGATCAAACTCCGTCGGAACCAACGATC 60
Db 1 TGGCGACTCTCTCTCTCTCATGTCACAGATCAAACTCCGTCGGAACCAACGATC 61

QY 61 TTCTGAGAAACACCCGCTCTCTCTCAATAGCTCCGGCGAGATCCGGACAGACACTGA 120
Db 62 TTCTGAGAAACACCCGCTCTCTCTCAATAGCTCCGGCGAGATCCGGACAGACACTGA 121

QY 121 AGAGTCGTGTCGGAACACAGGATCGAAGACAGTCAGCAAAACAGAAACCAACGAT 180
Db 122 AGAGTCGTGTCGGAACACAGGATCGAAGACAGTCAGCAAAACAGAAACCAACGAT 181

QY 181 TGAGAGGAATGGGTGTAGCAAAAGCTCCAGCTCAGAGCAATCGAAGAAAGAGCAAC 240
Db 182 TGAGAGGAATGGGTGTAGCAAAAGCTCCAGCTCAGAGCAATCGAAGAAAGAGCAAC 241

QY 241 TCGCCGCCGCCACAGTCGCGAGACACGTCATCAGTAGCATGCTCTAACAACGCTACCC 300
Db 242 TCGCCGCCGCCACAGTCGCGAGACACGTCATCAGTAGCATGCTCTAACAACGCTACCC 301

QY 301 GTTTACCCGTACCGGTAGACCCGGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGGA 360
QY 302 GTTTACCCGTACCGGTAGACCCGGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGGA 361

QY 361 GCAACAGGATCTATTGTGTGGAGTCGGGTCCGGTTCAGTTATGATCGACCCGGTTATT 420
Db 362 GCAACAGGATCTATTGTGTGGAGTCGGGTCCGGTTCAGTTATGATCGACCCGGTTATT 421

QY 421 CTCATCGGGTTTTGTGTGAGACCTCTCCCACTACTCATGCTCTCTCAATCTCAAAATC 480
Db 422 CTCATCGGGTTTTGTGTGAGACCTCTCCCACTACTCATGCTCTCTCAATCTCAAAATC 481

QY 481 CTCAAATGTTTACGCTTCTTCCAAATATCGCTGTGACACTTGTCTCA----- 528
Db 482 CTCAAATGTTTACGCTTCTTCCAAATATCGCTGTGACACTTGTCTCAAGGTTTGTGT 541

QY 529 ----- 528
Db 542 TTTTAAATCGTTTTCATCAACATGATGATATATATAGTTTTCGACTTTGAAAAAGTT 601

QY 529 ----- AGAAGAAACGTTTGGATGGTGGATCAGATAATG 561
Db 602 TTGATTTTATTTATGTAATAAACTCGAAGAAACGTTGGATGGTGGATCAGATAATG 561

QY 562 TAGTTCGATCCAAAGGTTGGATTTTCGAAATACAAATGATTCCTCTCCCGATGAACG 621

Db 662 TAGTTCGATCCAAAGGTTGGATTTTCGAAATACAAATGATTCCTCTCCCGATGAACG 721
QY 622 GCTAGCATCAGTATCTTCTCAATCAGATCATCATCAGAGGAGCCAGGTTTCTCTTATG 681
Db 722 GCTAGCATCAGTATCTTCTCAATCAGATCATCATCAGAGGAGCCAGGTTTCTCTTATG 781
QY 682 ATCATAGATCGCTAGAGCAGCTTCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTTATT 741
Db 782 ATCATAGATCGCTAGAGCAGCTTCACTTCTCTCTCTCTCTCTCTCTCTCTCTTATT 841
QY 742 TCAACGAGGCAACAAATCATAC----- 763
Db 842 TCAACGAGGCAACAAATCATACGCTACTAGTATAGTCCATTTTATTAACTCATATATA 901
QY 764 -----GGGACCAAT 772
Db 902 GGTATATATGATATATACCTGTTGATCTTATTGATTTAACTGGTGGGTTTGGGACCAAT 961
QY 773 GGAGGAATTTGGGAGCTACATCGAAGGAAACCTAGAAAATGATCAGAGGAGTGTGAAGGA 832
Db 962 GGAGGAATTTGGGAGCTACATCGAAGGAAACCTAGAAAATGATCAGAGGAGTGTGAAGGA 1021
QY 833 GTACGAGTTTTTCCGGGGAATATGTTGAAAGAGTTTTCAGTGGTGGCTACAACTGCTC 892
Db 1022 GTACGAGTTTTTCCGGGGAATATGTTGAAAGAGTTTTCAGTGGTGGCTACAACTGCTC 1081
QY 893 ACTCTGAGTGTGATTCAGTCCCTAATACCATTTGTTCTTCAAGCTTTAA 944
Db 1082 ACTCTGAGTGTGATTCAGTCCCTAATACCATTTGTTCTTCAAGCTTTAA 1133

RESULT 3
ATF27G19/c 82381 bp DNA linear PLN 01-JUN-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone F27G19 (BSSA project)
DEFINITION Arabidopsis thaliana
ACCESSION AL078467
VERSION AL078467.1 GI:4972065
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 82381)
AUTHORS Bevan, M., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A., Bancroft, I., Mewes, H.W., Mayer, K.F.X., Lemcke, K., Mannhaupt, G. and Schueller, C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 82381)
AUTHORS EU Arabidopsis sequencing, project.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES Location/Qualifiers
source 1. 82381
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
misc_feature 1. 30951
/note="overlap to BAC M4122; please refer to this entry for analysis and annotation"
misc_feature 5126..5143
/note="there is a mismatch in this homopolimeric stretch